DOCUMENT RESUME

ED 026 871 EM 007 107

By-Markle, David G.

The Development of the Bell System First Aid and Personal Safety Course; An Exercise in the Application of Empirical Methods to Instructional Systems Design. Final Report.

American Inst. for Research in Behavioral Sciences, Palo Alto, Calif.
Spons Agency-American Telephone and Telegraph, New York, N.Y.
Report No-AIR-E81-4-67-FR
Pub Date Apr 67
Note-39p.
EDRS Price MF-$0.25 HC-$2.05
Descriptors-First Aid, *Instructional Design, Instructional Films, *Material Development, Test Selection, *Time Factors (Learning)

The project objective was to develop a basic first aid course which would teach at least as much in 7-1/2 hours as a standard ten-hour Red Cross course. Student performance on empirically revised editions of test questions led to the final design of the course. Data gathered from tryouts of 16mm pilot footage were used to write scripts for the final 35mm color films, six accident vignettes which introduced lessons in accident prevention and emergency care. On one wide-range posttest, out of a possible 326 points untrained subjects achieved a mean score of 85, subjects trained in a standard course 145, and subjects trained in the new course 270. Time-worth estimations and the elimination of common knowledge material contributed to the efficiency of the instructional product. (TD)

ED 026871

AIR-E81-4/67-FR

FINAL REPORT:

THE DEVELOPMENT OF THE BELL SYSTEM FIRST AID AND PERSONAL SAFETY COURSE

An Exercise in the Application of Empirical Methods to Instructional System Design

David G. Markle

APRIL 1967

American Telephone and Telegraph Company

AMERICAN INSTITUTES FOR RESEARCH/PALO ALTO

ED 026871

AIR AMERICAN INSTITUTES FOR RESEARCH

P.O. BOX 1113, PALO ALTO, CALIFORNIA 94302
(LOCATION: 1791 ARASTRADERO ROAD)

AIR-E81-4/67-FR

FINAL REPORT:

THE DEVELOPMENT OF THE BELL SYSTEM FIRST AID AND PERSONAL SAFETY COURSE

An Exercise in the Application of Empirical Methods to Instructional Systems Design

David G. Markle

April 1967

American Telephone and Telegraph Company

EM007107

Persons wishing to obtain information on the course and its application may contact the

Office of the Safety Supervisor
American Telephone and Telegraph Company
195 Broadway
New York, New York 10007

FOREWORD

The practitioners of programed instruction have performed a great service to educational technology by their emphasis upon empirical tryout and revision of draft programs. Even if linear programed instruction in its present form should disappear from the face of the earth in the future, this important methodological contribution can and apparently will have permanent value for the designers of instruction.

The project reported here illustrates in detail the significance of the above statement. While the purpose of the project was a straightforward practical one, namely to produce better training in a shorter period of time, this practical goal, of importance to the sponsor, is far exceeded in importance by the methodological innovations which were followed in the conduct of this project. While programed instruction is now customarily tried out and revised once a draft is available for this purpose, this procedure by no means includes all of the empirical steps which can be taken in developing a new training program. As this report illustrates, the method of empirical development can be used not only to revise draft programs but also to revise draft films and other media. Furthermore, and of even more importance, the empirical method can be extended to determination of training objectives and to the determination of the specific developmental steps to be taken. It is therefore felt that this report is of general significance and will be of great interest to training engineers.

Leslie J. Briggs, PhD
Director
Instructional Methods Program
American Institutes for Research
Palo Alto

ACKNOWLEDGMENTS

The proposal which led to this project offered the sponsor several alternate methods of developing the final course. The alternate methods differed in the extensiveness of the empirical development techniques to be employed. First acknowledgment is due Mr. G. E. Bruggeman, then Outside Plant, Training and Organization Administrator, AT&T, who decided in the first place to adopt the most extensive plan. A different decision at that time would have prevented many of the more technically exciting aspects of the project from taking place.

Many people in AT&T and the member telephone companies have contributed to the project. These include supervisory personnel in New York, tryout coordinators and test subjects throughout the country, and Pacific Telephone and Telegraph Company employees who provided liason with AT&T and served as test subjects and film actors. Special mention is due Mr. W. R. Arehart, Staff Representative, AT&T, who as project supervisor patiently bore with us through the long months of data gathering, analysis, and revision, when little apparent progress was made, and Dr. H. O. Holt, of Bell Laboratories, who served as technical consultant and advisor.

The American National Red Cross provided generous subject matter consultation throughout the project. Mr. R. M. Oswald, Deputy Director, Safety Services, reviewed all test and instructional materials in every developmental phase, and personally supervised the final month of first aid skills filming. The integrity of the final product owes a great deal to his efforts.

Mr. Al Niggemeyer, of ALN Productions, Inc., San Francisco, directed all filming and film editing. He was responsible for the technical aspects of film production and, perhaps more important, must be given recognition for his patience with and understanding of empirical methods. These involve much different constraints than do conventional film-making methods.

Among AIR staff, I should like to thank Dr. Leslie J. Briggs, Director of the Instructional Methods Program, for his guidance and encouragement throughout the project. Dr. Lloyd Brooks provided the initial inspiration for the use of time data, made available his laboratory equipment for that purpose, and carried out the accident literature research. Mrs. Madalynne Chapman was responsible for the initial procedural skills research, prepared draft scripts, and coordinated all filming. Mr. John Clark supervised the automated testing, analysed the time data, and drafted and revised the majority of the printed instructional materials. Mrs. Sally Ford carried out the

decision level classification, supervised the scoring of the field tests, and administered and summarized the 16mm film tryouts.

Each of these members of the project staff made major contributions to the final product. I should like to thank each for his loyalty, creativity and hard work.

David G. Markle
Project Director

TABLE OF CONTENTS

Summary - 1

Project Objective- 1
Methods- 1
Results- 1

General Course Description- 2

Scope and Objectives of the Course - 2
Design of the Course - 2

Developmental Methods - 3

Subject Matter Analysis- 3
Empirical Specification of Objectives- - - - - - - - - - - - - - - - - - 7
Empirical Design of Instructional Materials- - - - - - - - - - - - - - 13
Combination of Criterion Questions and Films - - - - - - - - - - - - 16
Course Tryout, Version 1 - 19
Revision of Version 1- 22
Course Tryout, Version 2 - 24
Revision of Version 2- 26
Course Tryout, Version 3 - 27
Limitations of the Data- 29

Conclusions - 31

Relative Power of the Developmental Methods- - - - - - - - - - - - - - 31
Design Decision Strategies - 32

SUMMARY

Project Objective

The objective of the project was to develop a basic first aid course which would, in seven and one-half hours, produce results at least equivalent to those produced by standard first aid instruction taking ten hours. According to an agreement between AT&T and the American National Red Cross, Red Cross certificates have been awarded following completion of standard first aid instruction given in the Bell System. Thus the new course was expected to meet Red Cross certification requirements.

Methods

Student performance data were used to support as many course design decisions as possible. A set of test questions, defined as potential course objectives, was pretested on trained and untrained members of the student population to determine the actual objectives. A revised subset of these test questions, without additions, was then used as a first draft of the course. Student performance guided the successive approximation of the final course through gradual alteration of the questions and addition of needed instructional materials. The strategy followed throughout was to add instructional materials to the basic test question sequences only when the need was revealed through student tryouts. Response-time data and error data were used in this process.

Empirical methods were used to develop the motion picture components of the course, as well as the printed components. Data gathered from tryouts of brief segments of 16mm black and white pilot footage were used to develop scripts for final filming in 35mm color. The 35mm films were in turn tried out and revised on the basis of student performance.

Results

These instructional engineering methods have resulted in the attainment of the project objectives. In addition to the desired increase in efficiency as a function of decreased time, the new 7 1/2 hour course is far more effective than the 10 hour standard courses with which it has been compared. On one wide-range test used for comparisons, untrained subjects achieved a mean score of 85, subjects trained in standard first aid courses achieved a mean score of 145, while subjects trained in the new course achieved a mean score of 270, out of a possible maximum of 326 points. Similar results were obtained with other tests and other subjects.

GENERAL COURSE DESCRIPTION

The following description of the completed course is quoted from the course Instructor Outline:

SCOPE AND OBJECTIVES OF THE COURSE

This course is based upon the content of the American National Red Cross Standard First Aid Course, and thus shares the purpose set forth in the American National Red Cross First Aid Instructor's Manual:

> "The purpose of first aid training is to acquire knowledge and skills for the emergency care of the injured until a physician arrives, and to create an active interest in the prevention of accidents through elimination of the causes."

The general instructional objective of the course is that the trainees meet the American National Red Cross requirements for the Standard First Aid certificate. This includes performance of basic first aid skills, and knowledge of how and when to apply such skills . . .

*

DESIGN OF THE COURSE

The course is introduced by a series of six filmed accident vignettes which set the stage for first aid, pose questions on critical aspects of the accidents, and carry implicit accident prevention messages. Continuations of these vignettes introduce lessons later in the course. A filmed explanation of course procedure follows the introductory vignettes, and precedes the first filmed lesson.

Major topics which involve basic first aid skills are treated by a series of demonstration films and practice sessions. Skills are presented in brief filmed demonstrations. Practice sessions, in which trainees practice the procedures they have just observed, immediately follow the filmed demonstrations. The length of each demonstration film has been kept to a minimum, to ensure good performance in the practice sessions.[1]

Following a series of films and practice sessions on any one topic, trainees work through a workbook which tests them on material learned from the film and practice session series, and teaches them further details about that topic. Topics which do not involve new basic skills, such as FIRST AID FOR HEART ATTACK, are taught entirely in the workbooks.

In total, there are 20 film segments, 17 practice sessions, and 13 workbook lessons, grouped into four units, each approximately two hours long. The course can be administered in one working day . . .

*

---

[1] Tryouts of pilot footage revealed student performance errors similar to those found in serial learning experiments. Breaking filmed demonstrations into short segments and interspersing frequent practice sessions eliminated these effects.

DEVELOPMENTAL METHODS

Subject Matter Analysis

The final product of the subject matter analysis phase was a set of performance requirements, couched directly in test question form, for all content the project staff could identify as potentially belonging in a set of objectives for basic first aid. These questions were developed through the review of subject matter references, and the application of several schemes of logical analysis.

Criterion question generation. The first step was to generate every reasonable-appearing test question implied by the basic first aid sections of the American National Red Cross First Aid Manual. Passages of an obviously orientational nature, such as a definition of 'first aid', were not used to generate questions. Most of the questions were written in open-ended form, so they would simulate reasonably well the problems which might face a first aider.

Answers to some of the questions could not be determined from the text, even though these questions had been produced from the text. This difficulty was often caused by the mismatch between the general level of the text and the practical "What do you do in this situation?" level of the questions. Ambiguities in the text and alternate interpretations of the problem situations were resolved through consultation with the Red Cross, and through the subsequent analysis in which first aid was treated as a decision-making topic, rather than as one composed primarily of procedural skills.

Decision level analysis.[2] The decisions required of a first aider were classified into five different general types, or levels, according

[2] Geary A. Rummler and Albert W. Schrader III, of the Center for Programmed Learning for Business, University of Michigan, were major contributors to the decision-making analysis.

to when they would take place during a first aider's chain of reasoning about an accident situation or victim. Each level was cast in the form of a general question type, to simplify the classification of the questions which had been generated.

The first level involved basic skills and procedures, and thus produced questions of a "How do you . . .?" or "Describe how you would . . ." type. The higher levels, which involved decision making that takes place before basic skills are applied, included (2), determining what specific action to take, (3), identifying the injury or illness, (4), predicting what is likely to be wrong with a victim under specific conditions, and (5), preventing accidents. Each question which had been generated from the text was assigned to one of these levels of decision making, and by sub-category within each level. The sub-categories delineated variations in question type, e. g., discrimination, listing, etc. Questions on general knowledge which did not fit these levels were assigned to a sixth, "K" category.

Sorting the questions by decision-making category provided a natural means of checking for omissions. The questions for each first aid topic (e.g., care for wounds) were examined to see if questions of each level of decision making were included where appropriate for that topic. This was equivalent to making a matrix with first aid categories along one axis and level of decision making along the other, then examining each cell for omissions. Many questions were added as a result of this analysis.

Decision flow charting. The third stage of subject matter analysis was to design flow charts for 16 key first aid topics, much as one would design for a computer to follow. Analysis in terms of the strict, binary, yes-no kind of decision making process followed by a computer typically reveals weaknesses in apparently clear specifications. This application proved to be no different: our first draft charts revealed many unclear areas.

Once the major difficulties were resolved through consultation with the Red Cross, and the decision flow charts were completed, it became possible to examine criterion questions from another reference point. The first aider's decisions were distinguished from processes or actions by the flow-charting conventions used. Decision points, for example, involved such questions as "Is it a thermal or chemical burn?", while action boxes involved straightforward actions, such as "Jut the victim's jaw." The decision points on the charts were compared with the existing set of questions and, where necessary, were used to generate new questions which would require students to make discriminations. The action boxes on the charts were similarly used to evaluate existing questions and generate new questions which would require descriptions or demonstrations of procedures. They also provided a guide to topics which would require demonstration on motion picture film.

In addition to the three procedures described above--production of questions directly from the manual, decision-level analysis, and decision flow charting--a number of other sources were used in the

subject matter analysis phase. Research in the available accident literature led to the production of questions on accident prevention. Some of these questions concerned dangerous aspects of apparently innocuous situations, while others concerned the frequency, hence importance, of specific injuries. Additional draft questions were provided by the Red Cross.

Objectives summaries. The final set of approximately 500 questions was subdivided by first aid topic, such as "care for wounds", "artificial respiration", "heart attack", etc. General objectives statements were abstracted from the questions for each topic. These statements might have been produced first, had different procedures been followed. In this case, however, the statements were intended to imply only what was contained in the questions from which they were derived, and are best thought of as summaries. The questions themselves were the basis of the objectives specification.

It would be customary to consider the objectives specification task to be complete at this point, since the conventional procedure is to specify objectives on a logical, rational basis, before engaging in instructional materials development and in empirical tryout and revision procedures. In fact, however, the set of questions and their summary statements comprised only the potential objectives from which the course objectives were to be selected. If this set of questions had been used as the objectives, the resultant course would have been far longer and far more inclusive than existing standard first aid courses.

The first aid text alone, not counting the additional sources used to develop the questions, contains more than can be taught, in any strict sense, in 7 1/2 or 10 hours. The Red Cross instructor's manual explicitly tells the instructor not to attempt to teach all that is in the manual. He is told to teach critical, important, life-saving points. Students are expected to read the remainder outside of class. Whenever instructional engineering methods are brought to bear on existing course materials, much that is normally "covered" must be reevaluated. Incidental topics which are taken seriously in behavioral terms can expand a course considerably.

At this stage of the project, the materials were in hand to proceed to the determination of exactly which of the potential objectives, as spelled out in the questions and their summary statements, were to be included in the actual objectives. Empirical methods, rather than purely logical or rational methods, were brought to bear on this problem.

Empirical Specification of Objectives

In this phase the set of criterion questions was pretested on members of the intended student population and refined on the basis of the data so obtained. There were two basic types of refinement: Questions were revised to eliminate difficulties which kept them from measuring the behavior they were intended to measure, and questions were dropped from the set of potential objectives when they were shown by the data to represent material already known by the student population.

Specific problems faced by this project. The limit of 7 1/2 instructional hours made it impossible to specify exact course objectives beforehand, since the amount which can be taught is dependent to a certain extent on time available. The flexibility afforded by "self-pacing", or by homework assignments of variable length, was not available. The task was to adjust the objectives to meet the time requirements.

Several different strategies were used to accomplish this. The question, "How many of these objectives should be included?" was changed to the two questions: "Which objectives can be omitted because the students already can do those things?" and "How much of what is left can be taught in 7 1/2 hours?" The first of these is properly part of objectives specification, and can be answered to a fair level of confidence with pretesting. The second lies in the area of course design methods. Both are empirical questions, and both are critical to making instruction as efficient as possible.

An additional unknown contributed to the difficulty of saying exactly what should be included in basic first aid. The new course was intended to produce performance at least equal to that produced by standard first aid training. But at the inception of the project, detailed objective performance criteria did not exist for these courses. The natural solution to this problem lay also in testing.

Field testing. To answer the two questions, "What do the potential students of the new course know already?" and "What do standard

first aid courses teach?" the criterion questions were administered to Bell System employees who had and had not received first aid training.

It was not feasible to administer all questions to all subjects in the field testing, so shorter tests were produced. The questions were randomly distributed into four groups, then tabulated by level of decision making, within first aid topic. Excessive imbalances in topic distribution were eliminated by redistributing some questions. The resulting four tests covered different points, but were approximately equally weighted by topic, and were of equal length. A more desirable procedure would have been initially to distribute the questions randomly within level. It had been expected, however, that overall random assignment would have provided a more uniform distribution of items than was in fact obtained.

Because items ranged across many topics, and were scored only qualitatively in order to provide data on needed course content, measures of internal consistency were not computed. Alternate-form reliabilities were computed for those test forms which were used later for quantitative purposes, and are reported on pages 23, 26, and 29.

The tests were administered to approximately 800 subjects who were selected on an availability basis by the telephone company from four geographical areas. Three levels of standard first aid training were represented: no training, just completed training, and training within past five years. Some control subjects used in the later quantitative comparisons were randomly selected from these groups. (See pages 21, 22, and 29.)

Responses were tabulated on an item-by-item basis, so the error pattern of each item could be examined. Items which were seldom answered incorrectly were classed as candidates for omission from the potential objectives.[3] Consistent error patterns were taken to indicate incorrect common knowledge which would require special attention in the new course. Typical incorrect answers involved rubbing frostbite, loosening tourniquets frequently, using pulse rate as an indicator of stopped breathing, raising the feet of victims with head injuries, and removing auto accident victims from their autos immediately.

Laboratory testing on automated time recording equipment. Entirely separate testing was carried out with individual subjects[4] on automated time recording equipment. This testing, which was an integral part of both the objectives specification and the later instructional materials development, provided the means for incorporating response-time data as well as error data into the development process.

Typically, error data alone are available to the course designer. In most instructional programing systems, student errors on program and test items are taken to indicate revision requirements. The kinds

---

[3]Because the open-ended nature of the test questions permitted a wide range of "incorrectness", a strict numerical criterion, such as 5% errors or less, was not applied. In general, errors of omission, such as failing to mention "treat for shock" at the end of a procedure, were weighted lightly, while a single error which might result in injury or loss of life excluded an item from the potential omission category.

[4]Ninety-four subjects, including both trained and untrained, were provided by the Pacific Telephone and Telegraph Company for the automated testing in the objectives specification and materials development phases.

of errors often provide leads on what kind of revisions are needed. Correct answers are reassuring to the programer, but seldom useful for revision purposes. The addition of time data permits a much finer analysis of the function of the question and makes possible revisions aimed at meeting practical time limits.[5] In order to carry out this type of analysis, the length of time a subject spends on a question is compared with the length of time typically spent on questions which are similar with respect to formal type, reading time, and response-request type. Question-answering time which is considerably shorter than normal reading time for items of that type is evidence that the student was not responding to all elements of the question, whether or not he answered it correctly. An excessively long answering time indicates difficulty and/or inefficiency. Often, ambiguities are signaled more by response time than by errors, particularly on items which most students eventually figure out.

The automated time recording equipment was first used with the same four tests which were administered in the field testing. The questions were presented one at a time on 35mm filmstrip viewer, in the sequence in which they had been printed for the field testing. Confirmation of answers was not provided. Total read-and-answer time for each item was recorded on a printing counter, in seconds in early stages of the tryouts, and in smaller units as refinement progressed. Read-and-answer times were converted to z-scores for each item type,

---

[5]See Brooks, L. O. Note on revising instructional programs. *Psychological Reports*, 1967, 20, 117-118, and Brooks, L. O. Response time during instruction. *Perceptual Motor Skills*, 1967, in press.

to facilitate comparisons of item efficiency between item types, as well as within.[6] In addition to time and error data, subjects provided subjective item-by-item comments in individual interviews following the laboratory sessions.

Data obtained from this testing were used both to eliminate common knowledge items, and to modify retained items. Items which had become candidates for omission by virtue of few errors in the earlier field testing were omitted only if they produced zero error and non-extreme times in the automated testing. Items which were answered correctly but which required unusually long answering times were not omitted, as the long answer times were taken to indicate either a confusion relating to item design, or difficulty with that topic. Items which appeared defective were revised, while topic-related difficulties were noted as evidence of needed training. Subjective comments, as well as time and error data, were used as a basis for clarifying confusing items.[7] Following revision, all items were resubmitted to the testing process.

---

[6]Had these z-scores been used for statistically precise inferential purposes, an initial normalizing transformation of the skewed time scores would have been desirable. The practical purposes of these procedures did not require this, however. A time score exceeding the limits of $\pm 3\sigma$ was used in early stages of the project as a signal that an item deserved close scrutiny and possible revision. As items were refined, the limits were reduced first to $\pm 2\sigma$, then to $\pm 3\sigma/2$.

[7]Subjects' comments alone have been found to be quite useful in this process. See, for example, Hovland, Lumsdaine, & Sheffield, *Experiments on Mass Communication*. Princeton University Press: Princeton, 1949, p. 26, on "Qualitative pretesting of items".

Ambiguities and inefficiencies were revealed in a surprising number of the items. Identification of such problems has critical implications for the later stages of instructional materials design, because an undetected inefficient question will lead to the design of concommitantly inefficient instruction. The inefficient question will lead to teaching the student both the critical content and how to answer the unclear question, if not just the latter. When measured performance is equated with objectives, each content-related change in the measurement instrument is a detail change in the objectives. In this context, it would perhaps be better to distinguish between global, overall objectives, which may remain comparatively fixed throughout a project, and detailed, functional objectives, which are much less stable. Functional objectives remain variable until any instructional engineering task is completed.

Empirical Design of Instructional Materials

General strategy. Development of all instructional components of the course followed the same general maxim: "Do not add instructional materials until you have evidence the student needs them." The course was approximated successively, starting with the refined set of criterion items and no additional material whatever as the first draft.

Development of the basic criterion question sequence. The previously described tests were converted into first draft instructional sequences by resequencing items. Items were first grouped according to first aid topic within each test form. Within topic, they were sequenced by level of decision making. Basic skill questions, for

example, were placed before questions about which skill to use. Then, within level of analysis, items were grouped in order of increasing response time. This sequencing was carried out mechanically through edge punches in the cards on which the items were typed.

The four newly sequenced versions of the tests were presented on the automated equipment with answer frames alternated with the question frames. Trial subjects would read an item, write an answer, advance the filmstrip to the correct answer, check the answer, advance the filmstrip to the next item. The equipment recorded read-and-answer time and answer-checking time. As earlier, times were converted to z-scores for comparison purposes.

Presented in this manner, sequenced but without additional instructional materials, each test was a remote approximation to an instructional program. Students had the opportunity to check their own answers or to learn from the answers presented. But the major function of the sequenced questions at this point was to obtain data. Time and error data were used to modify individual items and to identify needed resequencing. Four major cycles of such trials and revisions were carried out, with essentially the same purpose of gradual data-based refinement. In later stages, the four separate semi-programs were combined to form the master criterion question sequence for the course. When sequences were combined, redundancies which were indicated unequivocally by time and error data were eliminated.

Up to this point, development of the questions had been separate from the development of the procedural skills films, which was carried out concurrently.

Film development methods. Early in the project we decided that the only topics which could be assigned to film treatment on an obvious *a priori* basis were those whose presentation and performance evaluation should involve demonstration of some sort. The "Demonstrate . . ." criterion items produced in the subject matter analysis phase were used as a starting point for the selection of such topics, which naturally included bandaging, splinting, artificial respiration, etc. Decisions about the inclusion of nonprocedural skills topics in the films were delayed until later, when empirical data on the requirements would be available.

As a first step, simple black and white 16mm films of basic first aid procedures were made. They were shot, whenever possible, at zero angle, so the procedures would be seen from the point of view of the first aider himself. The intent in this filming was simply to show how each procedure should be carried out. No special attempt was made to achieve continuity between scenes, nor to achieve artistic effects.

The films were broken down into short, unnarrated segments and mounted on 100 foot reels. The testing procedure was to project a single reel, then tell the trial students to do what they had just seen done on the film.[8] Their performance was observed and notes were made of errors and confusions. No instruction other than what was contained in the silent motion picture film was presented initially. The experimenter replied parsimonously to questions only after the student's difficulty with the footage had been tentatively identified. Consistent

---

[8]Sixty-nine trial subjects were provided by the Pacific Telephone and Telegraph Company for the film segment tryouts.

errors and explanations needed from the experimenter were used as indicators of needed reshooting and/or narration. In cases in which alternate versions of a procedure had been filmed, the more effective versions were identified. Reels dealing with related topics were tested in a variety of sequences to identify advantageous sequencing patterns.

After the more effective films had been selected, the clearly defective films reshot, and narration guidelines developed, the films were assembled into more molar units for tryouts in combination with the criterion question sequences. Short segments of film which had been tested separately in silent form were spliced together and narration was added where testing had indicated a need. White leader and continuous still pictures were used to fill in where the narration took longer than the related motion pictures. Conflicts between audio and visual channels were avoided by minimizing narration during filmed sequences of critical procedures, and by minimizing potentially distracting visual interest when the narrator had a critical explanation to make.

Combination of Criterion Questions and Films

Films and question sequences were first grouped together by topic. Both films and criterion questions were available at this point for topics which have both substantial procedural and decision making aspects. Bandaging had only films, since no nonprocedural questions on bandaging remained in the final set of questions. Nonprocedural

topics like heart attack had only a sequence of questions, since no films had as yet been produced on such topics.

In cases where both films and question sequences were available for a topic, the filmed demonstrations and coordinated practice sessions were placed first, followed by the printed question sequence. The length of an individual filmed demonstration, hence the number of skills to be practiced in an individual practice session, was determined through a compromise of conflicting requirements. Our general preference was to keep the film segments quite short. Student performance on any one task would clearly be benefited by the absence of other material intervening between the introductory demonstration and the practice session. Poor performance could be predicted on a topic which was presented at the beginning of a long series of demonstrations, if practice was delayed until the end of the series. On the other hand, frequent stopping of the film for practice sessions threatened administrative complications and time wasting.

In Version 1, the typical film demonstration session presented between one and three procedures, according to their complexity and interrelatedness. For example, techniques for the care of wounds were grouped into two sections. The first film segment treated direct pressure, elevation, and the pressure points, and was followed by a practice session for all three. This was followed by a film and subsequent practice session on the tourniquet. The question sequence on the control of bleeding followed the tourniquet practice session.

The sequences of questions were not changed from their criterion question sequence form as produced in the automated testing. Each sequence was, however, prefaced with a page or two of explanatory text on material which was clearly not covered by the films and practice sessions which now preceeded the sequence. For example, the explanatory text which was prefaced to the set of questions on Care for Wounds did not contain information on direct pressure, elevation, pressure points, or the tourniquet, as these were treated in the films and practice sessions. It did contain information on care for minor wounds, since this was not filmed. Set up in this manner, the questions functioned as a programed review and test on the films and on the brief printed text passages. Criterion question sequences for topics which did not have films were treated similarly. The sequences were unchanged, but were prefaced with brief introductory texts. The same general strategy of adding a bare minimum of material, as followed in other stages of the project, was followed here also.

A review unit was constructed for the end of the course by selecting critical questions from each lesson. High error frequency in earlier testing and potential life-saving value of the question were used as criteria for this selection.

It should be noted that at this stage the course was hardly "presentable". It had no introduction, no "motivational" material, and very little apparent continuity. All the refinement techniques applied up to this point had been concerned strictly with student performance, with no attention at all being paid to surface appearance.

Certainly no conventional film producer would have shown the film to a client as an example of his skill, nor could any conventional programing criteria have been used to evaluate the workbooks which contained the brief texts, the questions, and their answers. These materials bore about the same superficial relationship to the final course as the concrete supports, still encased in their forms, bear to a completed bridge.

Course Tryout, Version 1

Version 1 was administered in Atlanta, Georgia to ten trainees who had received no prior standard first aid training. The instructor was given a brief introduction to the course by project staff and Mr. Oswald, the subject matter expert, but he received no formal training in administering the course. Neither was he given a detailed instructor's guide for the practice sessions, as one goal of this tryout was to identify administrative difficulties and instructor's guide requirements. No unusual administrative difficulties were identified, other than the instructor's natural difficulty in changing abruptly from lecturing about first aid, as he normally did in his work, over to the much more limited and specialized tasks required of him by the new course. Project staff and Mr. Oswald were on hand to help out when necessary.

Total instructional time for Version 1 was approximately twelve hours, with the films taking two hours, practice sessions taking three, and workbooks taking seven. It was obviously hard work for the trainees, who commented freely that "this is sure harder than

high school." Mild complaints accompanied the distribution of the workbooks in the last few hours of the course. The films, although hardly exciting, were obviously looked forward to as respite from the workbooks, which could be described as an extreme form of "brute force" programing. The students were observed to adopt a style of working quite different from that which is normally observed with low error-rate programs. They would read a question, puzzle about its answer, make a guess at it (usually part of the complete answer), then study the correct answer given on the next page. Obviously the answers were not serving the occasional confirmation or answer-checking function they serve in conventional low-error programed instruction, but were often serving as initial instruction. That this worked, if inefficiently, is shown by the test results presented in Tables IA and IB.

TABLE IA

Test Form I Comparisons of No Instruction, Standard Course, and Version 1 of the New Course
Maximum Possible Score 351 Points

	No Instruction	Standard Course	Version 1
Course Length	0 hrs.	10 hrs.	12 hrs.
Mean Score	109	132	267
Standard Deviation	30	36	29
Lowest Score	41	39	219
Highest Score	149	244	325
Range of Scores	108	205	106
Number of Subjects	10	30	10

TABLE IB

Test Form II Comparisons of No Instruction, Standard Course, and Version 1 of the New Course
Maximum Possible Score 326 Points

	No Instruction	Standard Course	Version 1
Course Length	0 hrs.	10 hrs.	12 hrs.
Mean Score	85	145	229
Standard Deviation	40	42	28
Lowest Score	4	64	180
Highest Score	141	215	284
Range of Scores	137	151	104
Number of Subjects	10	22	10

The no-instruction control group and the group which received Version 1 of the new course were randomly selected from the same pool of newly hired employees. Several last minute substitutions in both groups by the telephone company upset the randomness, but did not introduce any identifiable biases. Both test Form I and test Form II were administered to the no-instruction control group and the Version 1 group, thus data on the same two samples are reported in Tables IA and IB in the No instruction and Version 1 columns. The control groups which had received standard first aid training were randomly sampled from the appropriate cells of the earlier administered field testing, and represent a pooling of results from numerous different standard first aid courses. Since only one test was administered to any single individual in the earlier field testing, different samples of 30 and 22 are reported for test Form I and test Form II. All testing was done immediately on completion of the course involved. The no-instruction control group was tested at the same time as the Version 1 group.

The product moment correlation between scores obtained on test Form I and test Form II by the subjects who received Version 1 was .82. No treatments or time interval intervened between administration of the two test forms. The correlation necessary for p = .01 is .765 (8 degrees of freedom).

Revision of Version 1

The major goal of the revision was to reduce instructional time from twelve hours to nearer the target of 7 1/2 hours without sacrificing the effectiveness of Version 1. Data obtained from the Atlanta tryout included observations of performance in the practice sessions, responses to questions in the workbooks, responses to test Forms I and II, and student comments.

Scripts for refilming all of the procedural skills films in 35mm color were prepared from the black and white films tested in Atlanta. Revisions suggested by the testing, aimed both at improved performance and at achieving a closer interdependence between films and workbooks, were incorporated into these scripts. In addition, scripts were prepared for nonprocedural topics which had been identified in the tryout as being troublesome or worthy of greater emphasis.

It was evident that explicit instructor guides would be required. Practice session guides were prepared for Version 2. These consisted of exact scripts for the instructor to read to the students during each practice session, and checklists on which he could check off skill points. The scripts were intended both to help the instructor organize the practice sessions, and to set up the situation so some testing of

conceptual material would take place. For example, the instructor was told to say, "The victim has a serious wound on his forearm. Do what you would do first to stop the bleeding" instead of, "Now demonstrate direct pressure and elevation for a serious wound on the victim's forearm."

The workbooks were extensively revised, with strong emphasis on time-saving techniques which would not sacrifice instructional quality. Consistent correct answers on a single point within the course, in the review section, and on the final tests, were used as evidence of redundancy, and some material on these points was eliminated, as was done earlier when the separate criterion question sequences were combined. Questions which were answered uniformly correctly, but which involved content judged to be of less than critical importance, were converted to statement form in order to eliminate time consuming response requests. This had the effect of restricting the material on which a student would repeatedly be tested to critical, high value items. Treatments of points on which consistent errors were made were of course modified to reduce the probability of these errors.

In addition to these data-based revisions, a series of accident vignettes was added to the course. Six situations in which first aid would be required were selected on the basis of frequency and/or instructional value. A filmed montage of these situations begins the course with a bit of drama, and sets the stage for first aid. Each situation is carried up to the point where the injury or illness occurs, then the scene is frozen and a critical question is posed. For example,

the automobile accident scene ends with a closeup of a bleeding victim and the question "What is the first step in controlling bleeding?" After the six situations and their questions have been presented, the film recycles through the question stills and gives brief answers. The critical "Hurry Cases of First Aid" and several other key problems are introduced at this point. Continuations of each of these situations, in which first aid is given to the victim, are used later in the course to introduce individual lessons.

Course Tryout, Version 2

The newly produced 35mm color films were combined with the practice session guides and the revised workbooks in a manner similar to Version 1, and tested on two trial classes in San Francisco. One tryout was informally administered by project staff, the second was formally administered by a Bell System first aid instructor. Preparation of the instructor for the formal tryout was brief, as it had been in the Atlanta tryout of Version 1. On this occasion, however, the intent of keeping his preparation brief was to enable us to evaluate the function of the instructor's practice session guide, without the confounding influence of additional instruction.

The only major difficulty observed in the formal tryout was with the instructor's guide. The checklists and instructor scripts had been printed separately, which made it difficult for the instructor to keep his place in both at once. This encouraged him to stop using one or the other and carry on *ad libitum*. Mechanical difficulties with

the materials have been eliminated by later revisions, but the tendency for the instructor to improvise has not been completely eliminated.

Student performance in the practice sessions was noticeably improved. Most of the procedural skills errors which had been frequent in the tryout of Version 1 were eliminated by the new film. Total instructional time was reduced from twelve hours to nine hours. This time saving resulted from the greater efficiency of the workbooks, and from the much closer interrelationship achieved between the films and the workbooks in Version 2. Minor changes in narration, which added virtually no time, had in some cases permitted us to eliminate some workbook material, and in others enabled us to avoid adding material.

Overall comparison data for Version 2 are presented in Tables IIIA and IIIB. The product moment correlation between scores on test Form I and test Form II for subjects who received Version 2 was .80. The correlation necessary for $p = .01$ is .798 (7 degrees of freedom).

An additional comparison is provided in Table II, where the Version 2 group is compared with a similar group of subjects who were given the Pacific Telephone Company's first aid course. This course consists of kinescopes, review by a live instructor, guided practice sessions, and frequent brief summary tests. These data are provided because the control group is clearly sampled from the same population as the Version 2 subjects, and the Pacific Course is closer in formal structure to Version 2 than is standard first aid instruction. The measurement instrument used was the Version 2 internal review unit,

administered without answers provided. It was used instead of one of the longer test forms used in the other comparisons because the control group was available for testing for only a brief period.

TABLE II

Pacific Course Control Group, Compared with Version 2 on Internal Review Unit, Max = 60

	Pacific Course	Version 2
Course Length	8 hrs.	9 hrs.[a]
Mean Score	41	54
Standard Deviation	5.7	3.05
Lowest Score	27	47
Highest Score	49	58
Range of Scores	22	11
Number of Subjects	19	9

$t = 7.56$; $df = 26$; $p < .001$

[a]9 hours includes the 1/2 hour review unit which was an internal part of Version 2, but which was given after the 8 hour Pacific course was completed.

Revision of Version 2

Observation of the practice sessions revealed the need for some renarration and resequencing of the procedural skills films and practice sessions. The treatment of mouth-to-mouth artificial respiration was converted from one to two film sequences, with the practice also divided from one into two sessions. This was done to eliminate confusion between several related procedures. No refilming was found necessary. Instructor scripts and checklists for the practice sessions were combined into one document to minimize the practical difficulties

observed in the tryouts, and a detailed time schedule was prepared on the basis of predicted time requirements for each component of the course.

Workbooks were revised according to the same strategies used previously, with further attention to time-saving changes in response requests. Interrelationships between films and workbooks were further refined. Questions on material taught in the preceding films and practice sessions were placed at the beginning of each printed lesson, with the pages of instructional text and their questions following. The pages of instructional text, which had previously been grouped together, were split into smaller sections and distributed more evenly throughout the question sequences. The resulting overall pattern for each workbook lesson was (1) initial question sequence on preceding films and practice sessions, (2) brief page of text on new material, (3) question sequence on new material, (4) brief page of text on more material, (5) question sequence, etc. The format of each question sequence was as before, with a question on one page, the answer on the next. In addition to these revisions, subject matter changes suggested by the Red Cross were incorporated.

Course Tryout, Version 3

The final version, as published, was tested in the Pacific Telephone and Telegraph Company. The class was administered by a Bell System instructor who had been given a fairly thorough preparation, and was completed within the prescribed seven and one-half instructional hours. Comparisons on test Forms I and II are presented in

Tables IIIA and IIIB. The correlation between scores on test Form I and test Form II was .96. The correlation necessary for p = .01 is .834 (6 degrees of freedom).

TABLE IIIA

Test Form I Comparisons of No Instruction, Standard Course, and Versions 1, 2, and 3 of the New Course
Maximum Possible Score 351 Points

	No Instruction	Standard Course	Version 1	Version 2	Version 3
Course Length	0 hrs.	10 hrs.	12 hrs.	9 hrs.	7.5 hrs.
Mean Score[a]	93	132	267	268	278
Standard Deviation	32	36	29	22	16
Lowest Score	27	39	219	234	251
Highest Score	158	244	325	298	302
Range of Scores	131	205	106	64	51
Number of Subjects	30[b]	30	10	9	8

[a]Comparison of means between Standard Course and Version 3 yields t = 16.04; df = 36; p <.001.

[b]The control groups reported in Tables I and III are the same, except for the addition of 20 cases to the no instruction Form I group. These additional cases were randomly sampled from the field testing, numerically scored, and added to provide a larger N, at the request of the sponsor.

TABLE IIIB

Test Form II Comparisons of No Instruction, Standard Course, and Versions 1, 2, and 3 of the New Course
Maximum Possible Score 326 Points

	No Instruction	Standard Course	Version 1	Version 2	Version 3
Course Length	0 hrs.	10 hrs.	12 hrs.	9 hrs.	7.5 hrs.
Mean Score[a]	85	145	229	256	270
Standard Deviation	40	42	28	23	9
Lowest Score	4	64	180	223	259
Highest Score	141	215	284	289	290
Range of Scores	137	151	104	66	31
Number of Subjects	10	22	10	9	8

[a]Comparison of means between Standard Course and Version 3 yields t = 12.88; df = 28; p <.001.

Limitations of the Data

The data reported above provide ample evidence that the new course produces improved performance on written tests given soon after instruction. Questions remain, however, about long term retention and real world performance, which cannot be answered within the limits of an initial development project.

Delayed retention. Delayed retention testing must eventually be included in the evaluation of any instructional system which is directed toward behavior which must be maintained without scheduled practice over time. Retention data will identify which elements of the material taught decay more and less quickly over time. This information can then be used both to revise the basic course further, and to design and schedule supplementary delayed review instruction.

Real world performance. When the objectives of a course involve behavior which occurs with high frequency in a natural setting, without danger to the participants, it is usually possible to validate final criterion tests against measured real world performance. Correlations between performance on criterion tests and on-the-job performance provides evidence on the task relevancy of the material taught in a course. When the events connected with the objectives occur with comparatively low frequency, and with high danger to some of the participants, test validation is much more difficult to achieve.

The Red Cross does have a great deal of accumulated experience with real world first aid problems, so there is good reason to believe that the course is related to lifesaving and injury-treating behavior

in emergency situations. But no direct test validation data can be provided. Since the course will be given to large numbers of people, however, there exists the possibility that data on first aid incidents which involve people who have been trained by the new course can be collected. These data may eventually be used to evaluate the short term instructional goals of the course.

A further note on objectives specification. The above discussion of delayed retention testing and real world performance evaluation suggests that the empirical specification of objectives can be extended considerably beyond the techniques which were used in this project. Clear cut distinctions between long term and short term objectives raise experimental questions about the functional effects of different short term objectives--questions which cannot be answered empirically until after an initial course development effort has been successful in producing the desired short term behavior.

CONCLUSIONS

Relative Power of the Developmental Methods

The comparative value of the methods used in this project cannot be assessed with the overall course comparison data which were obtained. A cost effectiveness analysis of the methods used would require an extensive experimental rather than developmental effort. It is nonetheless worthwhile to comment on some aspects of these methods.

Although the main innovative interest of the project lies in the wide range of empirical methods used, the importance of the initial analysis of first aid in decision-making terms must be emphasized. It is critical not that the analysis was in terms of decision making *per se*, but rather that the analysis systematically shifted emphasis from content to behavior, crosscutting existing classification categories. It appears that the discovery of an analysis scheme which crosscuts existing content-related classifications is of critical importance to projects of this sort.

Techniques for the elimination of common knowledge material were undoubtedly major contributors to the large differences in scores found between the control groups and groups taught by the new course. Standard first aid instruction contains a large amount of common knowledge material, thus wide range tests show small differences between untrained and conventionally trained students. Because so much instruction in most subject matters is hamstrung by the reiteration of already known material, this aspect of the empirical specification of objectives deserves wide general application.

The successive approximation of instructional materials, starting from criterion questions alone, eliminated the need for initial judgments about what instruction students would require--judgments which are typically made with inadequate information. This procedure also made it possible to exercise careful control over the material which eventually was included in the course. This control is particularly important when time constraints are involved, and when several media of instruction are to be integrated. Time data further facilitated this close scrutiny, calling attention to poorly functioning components which otherwise might not have been identified. Additionally, of course, the time data provided the means for achieving high efficiency through detailed time-worth estimations.

Design Decision Strategies

It is interesting to note that little attention was paid directly to the question of medium selection. Apart from the *a priori* decision to film the skills demonstrations, medium selection decisions were typically made for very small components of the course, in the context of specific data-based requirements and physical constraints. It appears that the highly detailed empirical work on the basic levels of objectives specification and instructional requirements determination eliminated the need for large scale decision making on the medium selection level. Design decisions made at any one level of development have surprisingly far reaching consequences at other levels--consequences which suggest that the sequence of decision making alone is well worth further attention, apart from questions of what kinds of evidence are to be used for each decision.